(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: 0 132 033
A1

(12) # EUROPEAN PATENT APPLICATION

(21) Application number: 84303740.9

(22) Date of filing: 04.06.84

(51) Int. Cl.⁴: H 01 L 21/82
H 01 L 27/10

(30) Priority: 19.07.83 US 516064

(43) Date of publication of application:
23.01.85 Bulletin 85/4

(84) Designated Contracting States:
AT DE FR GB IT NL SE

(71) Applicant: AMERICAN MICROSYSTEMS, INCORPORATED
3800 Homestead Road
Santa Clara, CA 95051(US)

(72) Inventor: Batra, Tarsaim L.
10316 Virginia Swan Place
Cupertino California 95014(US)

(74) Representative: Jones, Ian et al,
POLLAK MERCER & TENCH High Holborn House 52-54
High Holborn
London WC1V 6RY(GB)

(54) **Process of producing custom programmed read only memory.**

(57) A late mask programming process is provided for factory programmed ROMs or logic circuitry. MOS transistors functioning as ROM cells or in logic circuitry are fabricated by a standard MOS Process. Then, a thin stop layer of silicon nitride is provided over the transistors and followed by a layer of silicon dioxide. Programming is accomplished by applying a program mask and etching through the layers overlying the gate regions of selected transistors down to the silicon nitride stop layer. The silicon nitride stop layer prevents overetching and shorting of the gates. Then, ions are implanted underneath the gates of the selected MOS transistors to alter their threshold so, for example, as ROM cells they signify a different state than those cells whose transistor gates are not implanted with ions. The silicon nitride layer serves to stop the etch solution but permits the ions to pass through, penetrate the substrate and raise the thresholds of the selected transistors. The silicon dioxide layer stops the ions from being implanted into the nonselected transistors.

EP 0 132 033 A1

./...

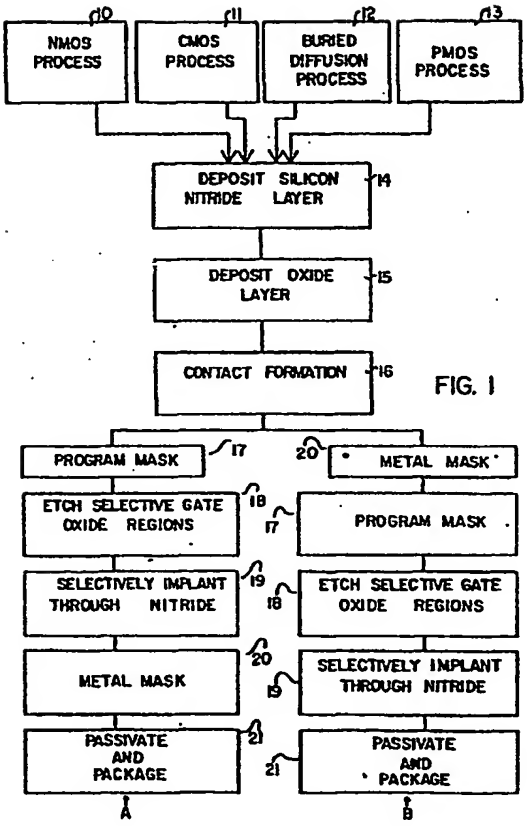

10
NMOS PROCESS
11
CMOS PROCESS
12
BURIED DIFFUSION PROCESS
13
PMOS PROCESS
DEPOSIT SILICON NITRIDE LAYER
14
DEPOSIT OXIDE LAYER
15
CONTACT FORMATION
16
PROGRAM MASK
17
ETCH SELECTIVE GATE OXIDE REGIONS
18
SELECTIVELY IMPLANT THROUGH NITRIDE
19
METAL MASK
20
PASSIVATE AND PACKAGE
21
A
20
METAL MASK
17
PROGRAM MASK
18
ETCH SELECTIVE GATE OXIDE REGIONS
19
SELECTIVELY IMPLANT THROUGH NITRIDE
20
PASSIVATE AND PACKAGE
21
B

FIG. 1

PROCESS OF PRODUCING CUSTOM PROGRAMMED READ ONLY MEMORY

DESCRIPTION

This invention relates to a process of producing a custom programmed read only memory (ROM).

In the semiconductor industry the fabrication of devices has progressed from discrete devices to devices integrating multiple devices on the same chip. Such integrated devices have progressed from small scale integration, to medium scale integration, to large scale integration and now to very large scale integration. These integrated devices typically have been designed by semiconductor manufacturers with the view to satisfying the requirements of a broad range of customers. Such customers could pick and choose from the offerings of all domestic and foreign commercial suppliers of standard semiconductor devices. For many applications, however, standard devices of this type have not been available to optimally meet the requirements of the particular application. For extremely high volume applications, large companies have been able to custom order parts or, in some cases, have set up in-house fabrication facilities to produce their own semiconductor devices to their own specifications. Small companies or low volume applications have not been able to advantageously utilize such custom fabrication. More recently, a number of companies have offered custom semiconductor devices based on so-called gate arrays which are alterable during various stages of the fabrication process in accordance with the design of the customer. These so-called silicon foundries are believed to provide a satisfactory answer to the system requirements of many users, particularly low-volume users. However, it is imperative that they do not introduce delays into the system design and development cycle.

In addition to the logic arrangements of gate arrays, programmable memory arrays such as read only memories (ROMs) are now in widespread use. The memory patterns in ROMs may also be fixed at the factory to reflect initial or reference data provided by the customer or may be programmed in the field by fusing links or by electrically programming EPROMs. In programming ROMs at the factory the requirement has been the relatively straightforward one of programming a predetermined pattern into the available cells of the memory during fabrication. The manufacturer of ROMs, by some means, incorporates into its process a bit map provided by the customer which will determine the state (ones or zeros) of each bit in the ROM.

In the programming of ROMs at the factory, the conventional approaches have involved process steps at intermediate stages of the process sequence. Thus, once a customer has placed an order, it has always been necessary to carry out the programming step and then complete the processing of the wafers through all the rest of the processing steps. This necessity has resulted in significant turnaround time from customer order to availability of prototype. The standard ROM fabrication technologies have required that the ROM cell, a single MOS transistor, be programmed to its zero-logic state by various techniques, techniques which have occurred early in the process sequence or which have unduly enlarged the cell size. For example, the oxide is sometimes made thicker under certain ones of the transistors, thereby increasing the threshold voltage of the transistor. Or, in the contact mask, the zero-state may be obtained by not making contacts to those gates requiring a zero-state. And in one process gates are not formed over selected prospective MOS transistors. See C. K. Kuo, "Method of Making a Metal Programmable MOS Read Only Memory Device," U.S. Patent No. 4,384,399; see also the references and discussion therein. In some processes the zero-state is programmed in the metal mask by not making an electrical connection with the gate and in still

other processes the sources or drains are offet from the gate for selected transistors which are to have a zero state; see R.S. Countryman, et a., "Method of Programming ROM by Offset Masking of Selected gates, U.S. Patent No. 4,380,866. See, e.g., C.K. Kuo, "Post-Metal Programmable MOS Read Only Memory", U.S. Patent No. 4,390,971. And in still other MOS processes, the programming of ROM cells occurs by implanting ions through polysilicon layers in selected transistors to increase the threshold to render them nonconductive and produce the zero state. With all of the above programming techniques, the programming occurs while the devices are being defined or a penalty of increased cell size is incurred.

The invention accordingly provides a process of producing a custom programmed read-only memory (ROM) in a silicon wafer, the process comprising the steps of:

fabricating an array of metal oxide silicon (MOS) transistors in a silicon wafer up to the stage of contact formation;

depositing a layer of silicon nitride over the transistors;

depositing a layer of silicon dioxide over the layer of silicon nitride;

applying to the wafer a program mask containing a pattern which represents a map of selected transistors in the array of transistors of which the threshold voltages are to be altered whereby the gate regions above the selected transistors are exposed;

etching the layers of silicon dioxide above the selected gate regions down to the silicon nitride layer; and

implanting ions through the silicon nitride layer into the silicon regions underneath the gates to alter the threshold voltages of the selected transistors, the silicon dioxide layer serving to stop the passage of the ions.

Definition of contacts to active device regions can precede the etching through the layers overlying the gate regions of the selected transistors down to the silicon nitride layer, which prevents over-etching and shorting of the gates of the selected transistors, by stopping the etch solution but permits the ions to pass through to penetrate the substate. Alternatively, the programming and all subsequent steps are accomplished after the metal mask step.

The present invention can thus provide a process of producing ROMs wherein ROM programming occurs late in the process. The process does not add to cell size.

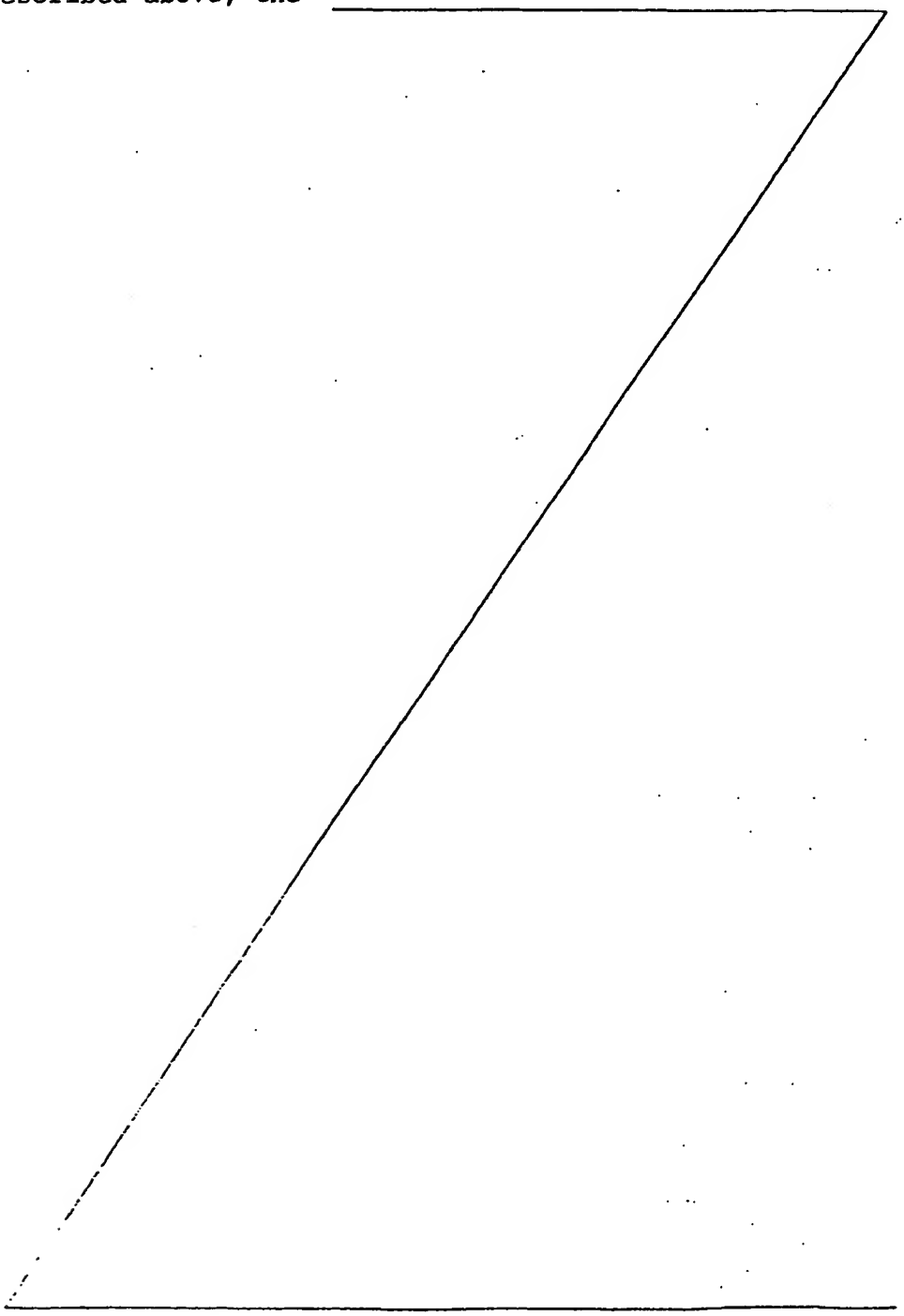
For a more complete understanding of the present invention, reference may be had to the following illustrative description and accompanying drawings, in which:

Figure 1 is a block diagram showing alternative fabrication technologies which may use at the end of the process sequence, a process in accordance with the present invention; and

Figures 2a-21 are a series of cross-sectional views of an MOS transistor being formed by an NMOS process (Figs. 2a-2g) and then being programmed (Figs. 2h-21) in accordance with a process in accordance with the present invention.

In producing custom integrated circuits, the overall process typically has four major stages. The diffusion or implantation stage defines and produces those areas such as sources or drains that are to serve as active electrical regions. The metal layer design stage interconnects the active electrical regions by conductive lines of metal or doped polysilicon to interconnect individual transistors (or cells) into logic or memory arrangements. The wafer personalization stage is where, in accordance with a customer's design, particular logic building blocks are selected and interconnected or where,

if not previously accomplished, ROM cells are programmed in accordance with the bit map of a customer. Then, in the final stage the device is packaged and tested. As described above, the



conventional approach to programming ROM cells is to carry out the programming in stages 1 to 3 and typically early in the overall process sequence. With the several techniques described above which occur late in the process sequence, a penalty in increased cell size is incurred.

As described previously, the conventional techniques for factory programming ROMs often include process steps which are carried out early in the process sequence. For example, the programming of the field oxide by providing thick oxide underneath certain transistors occurs as early as the field oxide definition step shown in Figure 2c. The omission of a gate over a particular prospective transistor could occur as early as step 2f. Selectively contacting the gates is a programming technique which occurs in the process at about the steps indicated in Fig. 2j, and, the omission of a metal interconnect through the metal mask occurs at about the step indicated in Fig. 2k. Either these steps occur relatively early in the process sequence or they result in a bigger cell size. Ideally, for the fast turnaround of a ROM to be custom programmed, a custom IC company would have wafers on the fully processed to the point of programming. With the late mask programming technique of the present invention, the setting of the state of a particular MOS transistor or ROM cell does not occur until the stage of the process indicated by Figure 2k. Cell size is not increased; only the threshold level of selected cells is altered. Wafers can be stored with all processing completed up to the stage of Figure 2k and then rapidly completed with the last several steps up to final packaging. This permits extremely fast turnaround for customers once they supply a bit map for programming of custom ROMs. Turnaround time can thus be measured in two to three days rather than in two to three weeks.

The process sequences for various embodiments of the present invention in the context of programming ROMs may be seen in the block diagram of Figure 1. Any conventional

MOS process for ROMs may be utilized to define the individual transistors. Thus, NMOS process 10, CMOS process 11, buried diffusion process 12 or PMOS process 13 may be utilized up to the stage of definition of the transistors. At this time in the personalization, a stop layer of silicon nitride is applied in step 14. This thin layer of about 100 angstroms to about 1000 angstroms serves later to prevent over-etching of the oxide in selective etch step 18, as described in detail subsequently. Next, an oxide layer is applied by conventional vapor deposition techniques to a thickness of approximately 10,000Å. Then the contacts are formed in step 16 to the individual sources and drains and to the poly interconnects and to gates. In embodiment A at this late time in processing a program mask with the bit map for programming the ROM is applied in step 17 to open up the oxide above the gates of those MOS transistors which are to be programmed to the zero state. The oxide above the gates of these transistors is etched in step 18, down to the thin silicon nitride layer applied previously in step 14. Preferrably, an etch is used which preferentially etches silicon dioxide over silicon nitride. The thin nitride layer, as described in detail subsequently, prevents over-etching so that there is no shorting of the gates to the source, drain or substrate regions. Then an ion implantation step 19 is carried out to raise the thresholds of those MOS transistors whose gates are exposed. The oxide remaining over the gates of the other MOS transistors prevents them from being affected during the ion implantation. The devices are then processed through standard processing techniques including metal evaporation and metal masking in step 20 and passivating and packaging in step 21. In embodiment B the metal evaporation and metal mask step 20 is carried out immediately after contact formation step 16. Thereafter, the program mask with the bit map for programming the ROM is applied in step 17 to open up the oxide above the gates of those MOS transistors which are to be programmed to the

zero state. The oxide above the gates of these transistors is then etched in step 18, down to the thin silicon nitride layer applied previously in step 14. In embodiment B, the thin nitride layer, as with embodiment A, prevents overetching so that there is no shorting of the gates to the source, drain or substrate regions. Then the ion implantation step 19 is carried out to raise the thresholds of those MOS transistors whose gates are exposed. The oxide remaining over the gates of the other MOS transistors prevents them from being affected during the ion implanatation. The devices are then passivated and packaged in step 21.

To fully examine the process sequence of embodiment A of the present invention and to particularly point out the stages at which programming is done in the prior art as compared with the process of the present invention, reference should be made to the process sequence shown in Figs. 2a-2l. As discussed above, the portions of this specific process through definition of the transistors may vary with other embodiments but will remain the same for the application of the stop layer of silicon nitride and the programming steps. The process sequence shown for defining the transistors is an NMOS process. In Fig. 2a, a silicon wafer 24 has a thin layer 26 of thermal oxide grown on its surface. A 700Å layer of silicon nitride 25 is deposited on the surface of silicon oxide 26. In Figure 2b the lateral expanse of a single MOS transistor is shown underneath resist region 27 and between field oxide areas to be formed. The transistor is exemplary of each transistor in a standard read-only memory array. The nitride layer 25 is etched everywhere except under the resist. Subsequently, a field implant 28 of 5×10^{12} atoms per square centimeter is introduced into the silicon substrate 24 through the exposed oxide layer 25 at an implantation energy of 100 kilovolts. Field oxide 29 is then thermally grown as shown in Fig. 2c. The field implant 28 is slightly diffused outwardly as a result of the high-temperature process.

An implantation of boron at $4x10^{11}$ atoms per square centimeter is carried out at an implantation energy of 50 kilovolts as shown in Fig. 2d to produce the gate implant region 19 and to provide the threshold voltage of the gate, typically about one volt. The implantation levels will vary as device requirements dictate. In a preferred embodiment before the gate implant is carried out, the initial oxide and nitride layers are stripped and a thin gate oxide is intentionally grown. Next, as shown in Fig. 2e, a section 31 of the surface of the silicon wafer 24 is opened up above the drain source region to be formed. The oxide is removed by chemical or dry plasma etching while a poly contact mask is in place. A layer 30 of highly doped polycrystalline silicon is then applied to contact the drain region to be formed. The polysilicon is defined in Fig. 2f to leave a gate region 32 and an interconnection 33 which makes electrical contact with the drain region 35 which has not been formed.

The final step in the definition of the MOS transistor is accomplished by the implantation shown in Fig. 2g. In this embodiment, $8x10^{15}$ atoms per square centimeter of arsenic is implanted at an implantation energy of 75 kilovolts. These ions are driven through the thin oxide over the source region 34 and are also driven into a drain region 35. In order to drive the arsenic further into the silicon substrate 24, a high-temperature thermal drive-in diffusion is carried out at a temperature on the order of 1050°c for about 15-20 minutes. This thermal drive-in produces source region 34' and drain region 35'. As a consequence of the thermal drive-in, the field oxide 29 is slightly increased in size and oxide layer 40 grows over gate 32 and over the exposed surface of drain region 35', source region 34', and polysilicon interconnect 33. This oxide region 40 will remain over the gate for the duration of processing. The MOS transistor illustrated in this process sequence is now fully formed and would be operational upon the application of gate, source and drain

contacts and the application of appropriate signals. It is at this late stage in the process sequence that the setting of the transistor to a zero state is accomplished. In accordance with the process of the present invention, a very thin layer 34 of silicon nitride is then applied to the entire surface. Preferrably, this layer has a thickness of 100-1000Å. The layer is thick enough to prevent over-etching by an etch which preferentially etches silicon dioxide over silicon nitride, such as dilute hydrofluoric acide, but is thin enough to permit ions to be implanted therethrough as described subsequently. Thereafter, as shown in Fig. 2i, a layer 36 of deposited oxide (called PVX) is applied over the entire structure. Typically, this will be deposited by chemical vapor deposition and have a thickness of about 10,000Å. The PVX serves to insulate underlying layers from overlying metal lines and stops ions from being implanted into transistors whose thresholds are not to be altered. As shown in Figure 2j, a contact mask is now used to open up access to the source 34', the drain 35', the polysilicon 33 and all other regions to which electrical contacts are to be made. The wafers are then processed through source and drain metallization to define interconnects to sources, drains, gates and polysilicon lines. Note defined interconnects 37 and 38 in Fig. 2k.

Up to this point in processing, no programming of ROM cells has occurred. Any logic transistors on the chip have been formed and indiscriminately interconnected and all ROM cells are formed and potentially operational. Subsequently, by programming, the transistors are rendered disfunctional in certain ROM cells or in certain portions of the logic circuitry by raising their thresholds so high, typically higher than 5 volts, so that they are will not function in the normal operation of the circuit. When a particular customer provides a bit map or specifies

logic circuitry, stored wafers may be taken out and processed in accordance with the bit map or circuit logic. The setting of an individual MOS transistor to a zero state is accomplished by masking the PVX oxide in accordance with the bit map as layed out on a program mask (step 17 in Fig. 1) and then etching through the PVX oxide down to the silicon nitride stop layer. Then, in one embodiment, to increase the thresholds of the selected transistors approximately 25×10^{13} atoms of boron per square centimeter is implanted at an energy of about 170 kilovolts. In another embodiment a species of opposite conductivity type is implanted to lower the threshold. What is required is that one set of transistors have one threshold level and another set have a different threshold level with the two levels being sufficiently separated to permit reliable differential detection.

In accordance with the process of present invention, the presence of silicon nitride layer 34 has allowed the etching over the selected gate regions to occur without over-etching and shorting of the gate to the source or drain regions or to the substrate. The silicon nitride acts as an etch stop once the etch has penetrated the overlying PVX and hence the etchant will not reach and attack the underlying layers. In other processes where implantation is used to alter threshold levels, without the presence of the nitride layer 34, any attempt to open up the areas of the gate region for implantation would have potentially resulted in the shorting of the gate. The implanted ions pass through the opened up region of the thin nitride layer, penetrates the gate region and enter the silicon under the gate to thereby raise the threshold to above five volts. The ions are stopped by the PVX elsewhere. In accordance with the preferred embodiment of the process of this invention, this programming occurs after contact formation (step 16, Fig. 1). After this late process step the ROM array is now fully programmed so that only a few steps yet remain such as

passivation, pad mask, scribing and packaging of the completed custom ROM; if logic circuitry is also on-board, the logic pattern will also have been established by raising the thresholds of selected transistors and effectively removing them from the circuits. After the programming, a passivating layer of silicon nitride 39 is applied over the whole surface. The integrated circuits containing the individual ROMs of Fig. 2ℓ are then scribed and packaged and provided to the customer. The completed integrated circuit will contain certain ROM cells which are operational transistors and other ROM cells whose threshold have been raised and are effectively disfunctional in normal operation.

In embodiment B of Figure 1, each of the individual steps of metal mask 20, program mask 17, selective gate etch 18, selective implant 19 and passivate and package 21 are carried out as described above for embodiment A. The program mask step 17 is carried out one step later than in embodiment A thereby permitting the metal mask step 20 to be carried out earlier in the process and slightly shortening the turnaround time from receipt of the customer's map to availability of prototype. The presence of the silicon nitride stop etch layer applied by step 14 is the key in both embodiments B and A to allowing the implantation step to be used effectively so late in the processing sequence.

CLAIMS

1. A process of producing a custom programmed read-only memory (ROM) in a silicon wafer, the process comprising the steps of:

fabricating an array of metal oxide silicon (MOS) transistors in a silicon wafer up to the stage of contact formation;

depositing a layer of silicon nitride over the transistors;

depositing a layer of silicon dioxide over the layer of silicon nitride;

applying to the wafer a program mask containing a pattern which represents a map of selected transistors in the array of transistors of which the threshold voltages are to be altered whereby the gate regions above the selected transistors are exposed;

etching the layers of silicon dioxide above the selected gate regions down to the silicon nitride layer; and

implanting ions through the silicon nitride layer into the silicon regions underneath the gates to alter the threshold voltages of the selected transistors, the silicon dioxide layer serving to stop the passage of the ions.

2. A process as claimed in claim 1 wherein the threshold voltages of the selected transistors are lowered by the step of implanting ions.

3. A process as claimed in claim 1 wherein the threshold voltages of the selected transistors are raised by the step of implanting ions.

4. A process as claimed in claim 1, 2 or 3 wherein the array of MOS transistors is fabricated by a NMOS process.

5. A process as claimed in claim 1, 2 or 3 wherein the array of MOS transistors is fabricated by

a CMOS process.

6. A process as claimed in claim 1, 2 or 3 wherein the array of MOS transistors is fabricated by a buried diffusion process.

7. A process as claimed in claim 1, 2 or 3 wherein the array of MOS transistors is by a PMOS process.

8. A process as claimed in any preceding claim wherein each selected transistor of the MOS transistors comprises a ROM cell.

9. A process as claimed in any one of claims 1 to 7 wherein at least one of the selected transistors of the array of MOS transistors is incorporated in a logic circuit.

10. A process as claimed in any preceding claim wherein the layer of silicon nitride is deposited in a thickness in the range of 100Å to 1000 Å.

11. A process as claimed in any preceding claim wherein the layer of silicon dioxide is deposited in a thickness in the range of about 7,000 Å to about 10,000 Å.

12. A process as claimed in any preceding claim wherein contacts to the active regions in the transistors are formed through the silicon nitride layer after the step of depositing a layer of silicon dioxide.

13. A process as claimed in any preceding claim having, after the step of implanting ions, the steps of:

applying a metal mask to the wafer and forming metal interconnections between previously defined contacts to active device regions; and

applying a layer of passivating material to the wafer.

14. A process as claimed in any one of claims 1 to 11 having, before the step of applying a program mask to the wafer and after the step of depositing a layer of silicon dioxide, the steps of:

forming contacts to the active regions in the transistors through the silicon nitride; and

applying a mask to the wafer and forming metal interconnections between the contacts to the active regions.

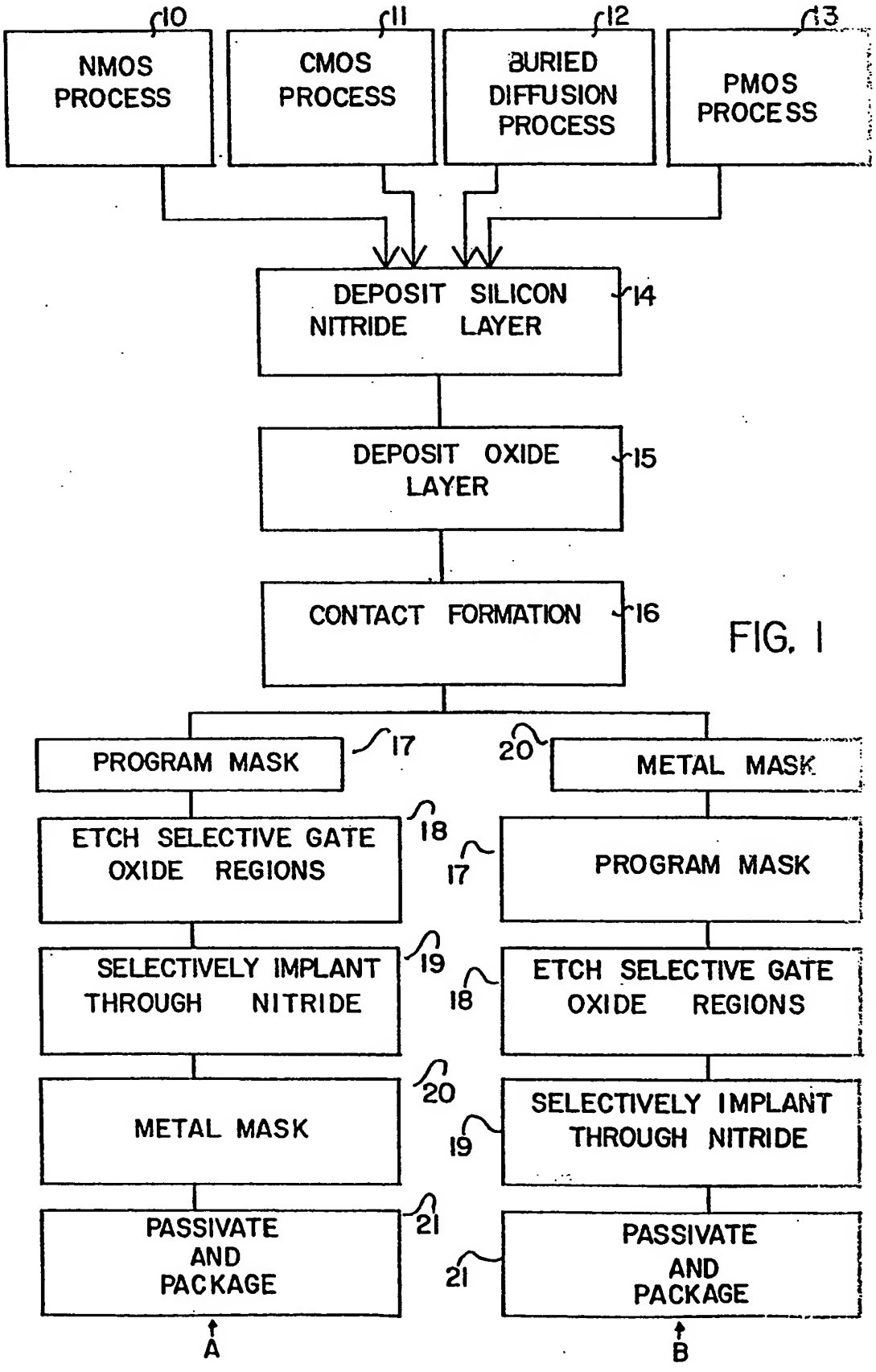

10
NMOS PROCESS
11
CMOS PROCESS
12
BURIED DIFFUSION PROCESS
13
PMOS PROCESS
DEPOSIT SILICON NITRIDE LAYER
14
DEPOSIT OXIDE LAYER
15
CONTACT FORMATION
16
PROGRAM MASK
17
ETCH SELECTIVE GATE OXIDE REGIONS
18
SELECTIVELY IMPLANT THROUGH NITRIDE
19
METAL MASK
20
PASSIVATE AND PACKAGE
21
A
20
METAL MASK
17
PROGRAM MASK
18
ETCH SELECTIVE GATE OXIDE REGIONS
19
SELECTIVELY IMPLANT THROUGH NITRIDE
21
PASSIVATE AND PACKAGE
B

FIG. 1

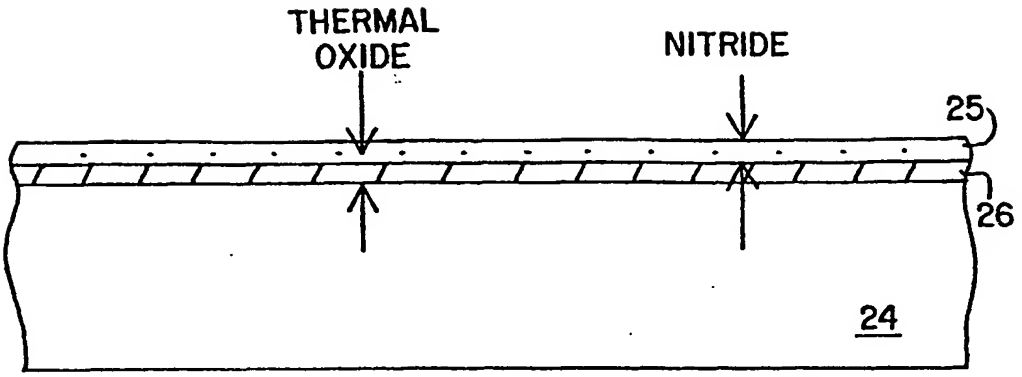


FIG. 2a

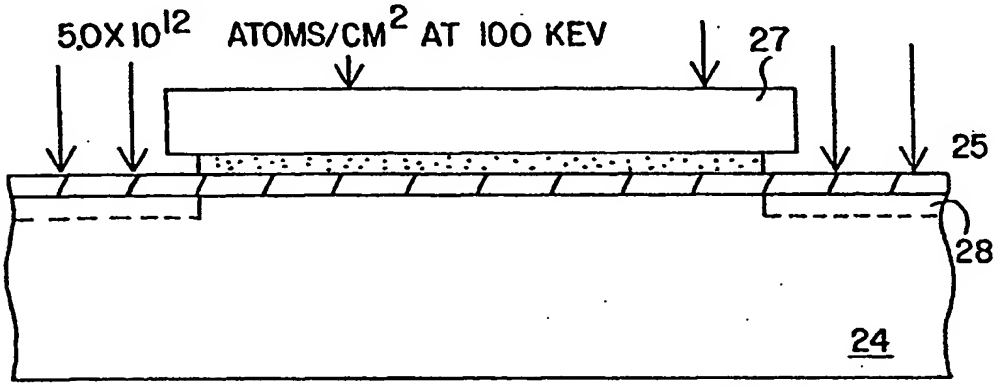


FIG. 2b

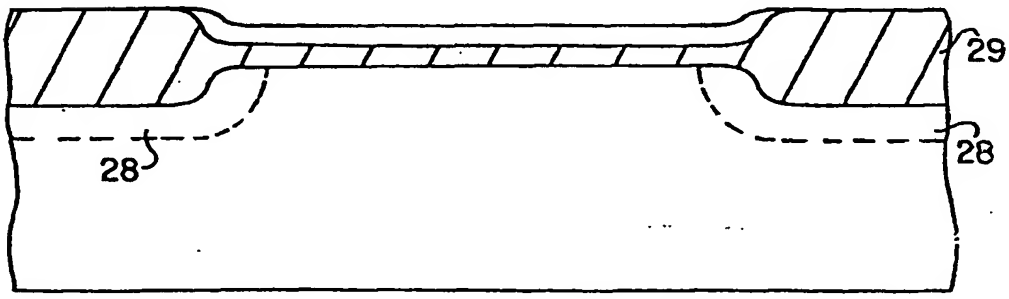


FIG. 2c

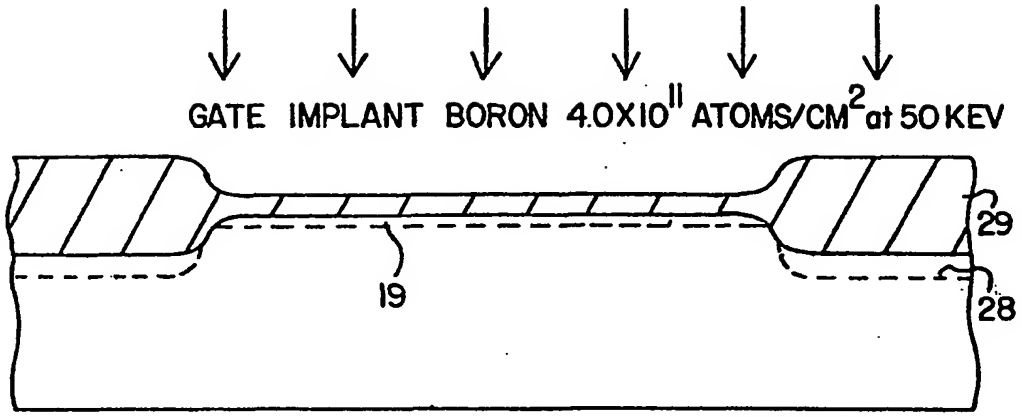


FIG. 2d

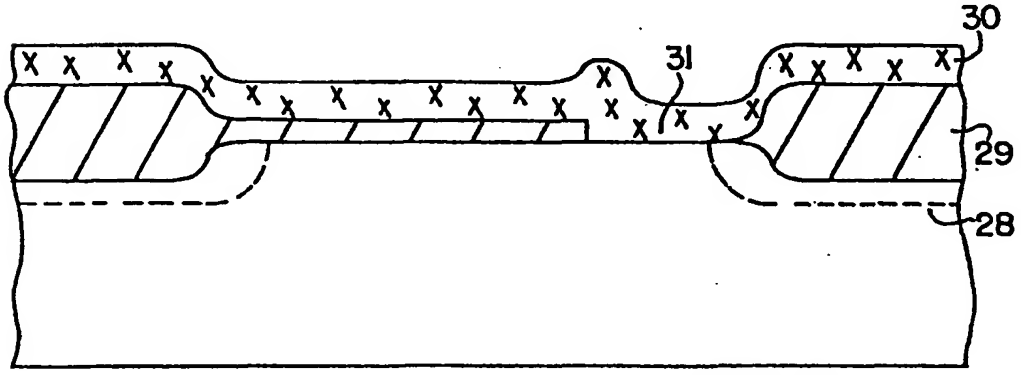


FIG. 2e

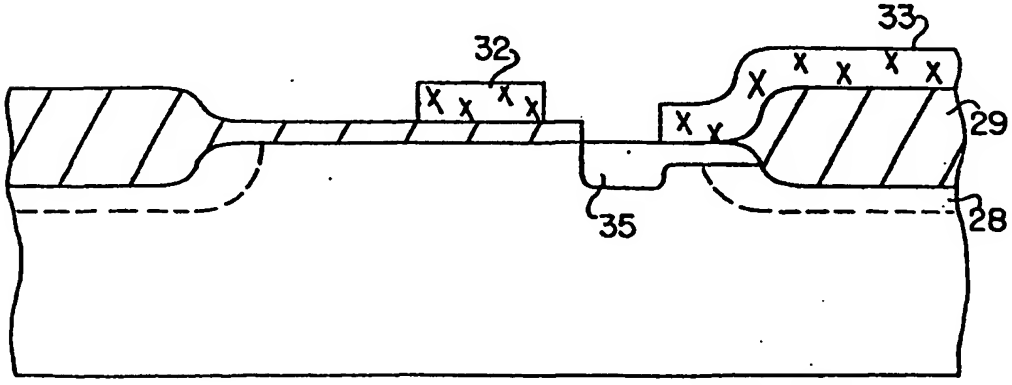


FIG. 2f

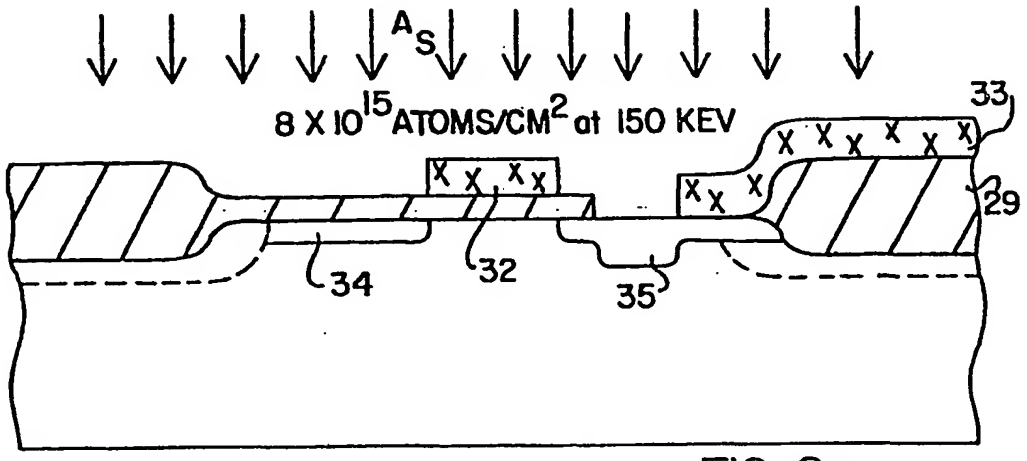


FIG. 2g

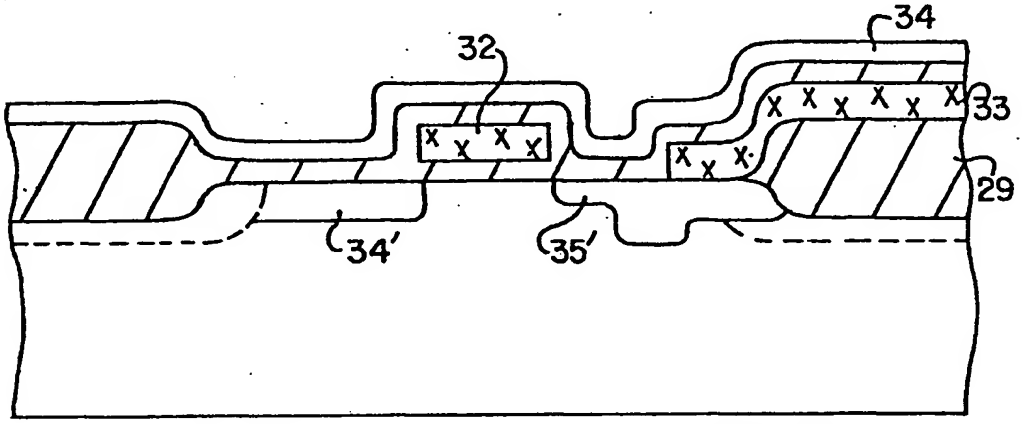


FIG. 2h

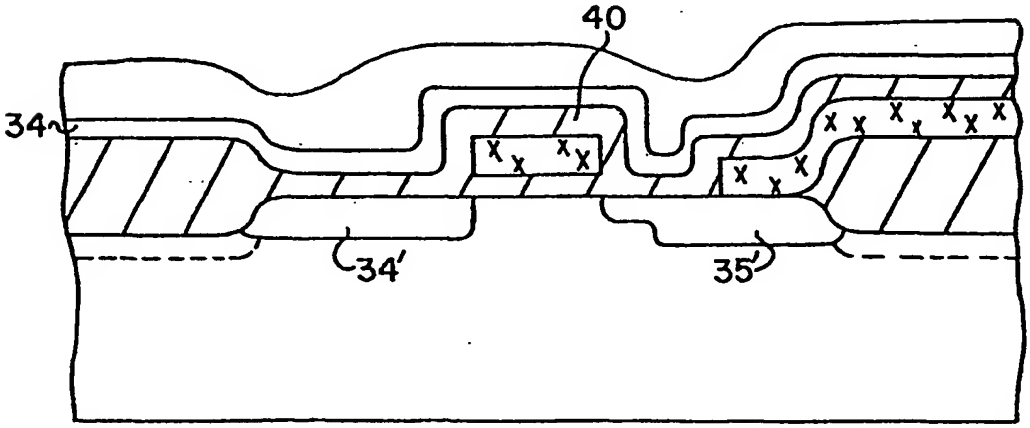


FIG. 2i

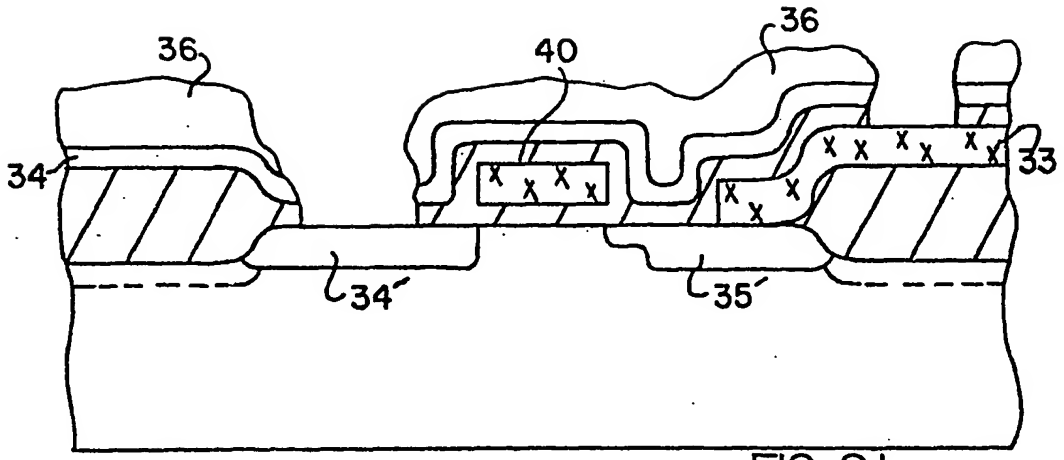


FIG. 2j

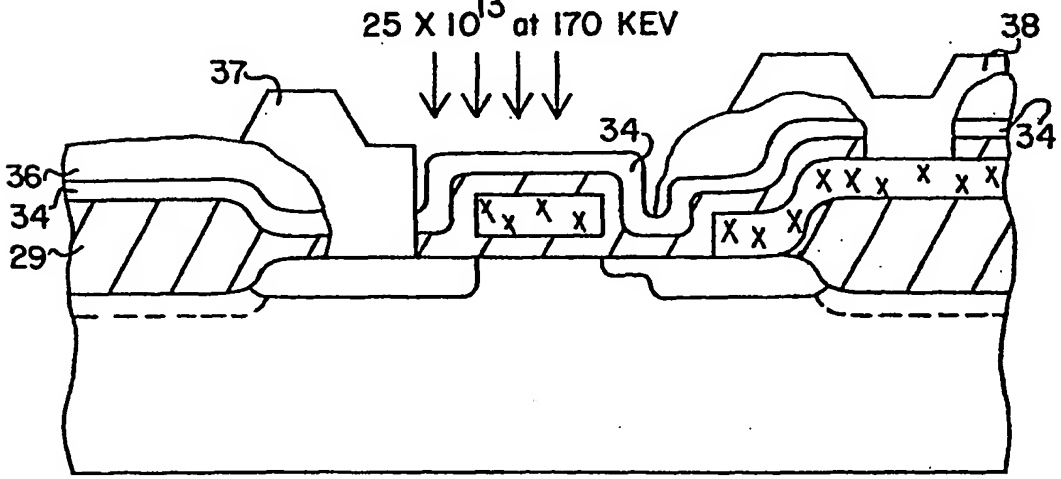


FIG. 2k

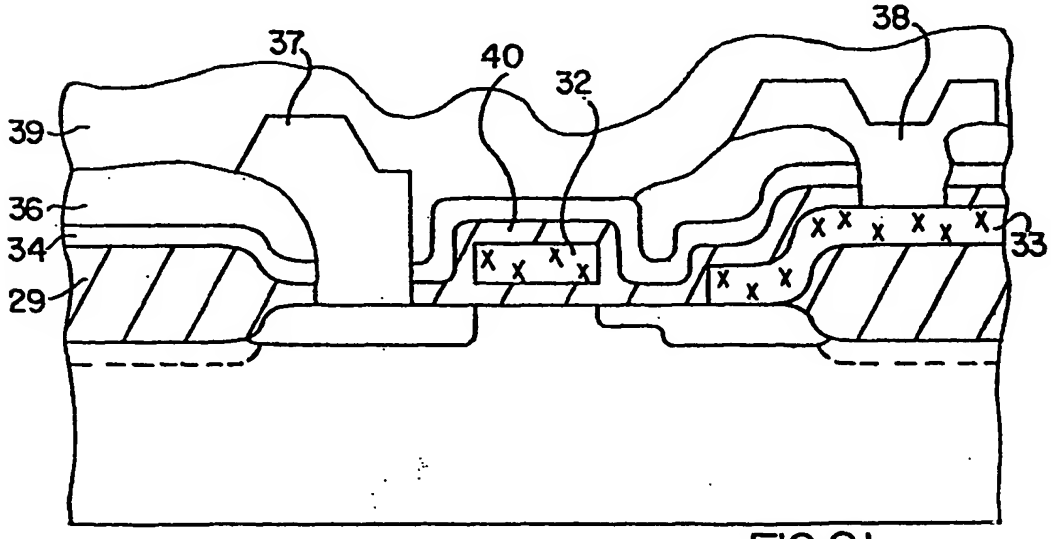


FIG. 2l

European Patent Office

EUROPEAN SEARCH REPORT

0132033
Application number

EP 84 30 3740

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl. ³)
X	US-A-4 295 209 (W.B. DONLEY) * Claims 1,2; column 5, line 62 - column 6, line 61; figures 6-11 *	1-3,8, 10,13	H 01 L 21/82 H 01 L 27/10
X	US-A-4 364 167 (W.B. DONLEY) * Claims 1-3; column 5, line 63 - column 6, line 62; figures 6-11 *	1-3,8, 10,13	
A	US-A-4 273 303 (P.K.CHATTERJEE et al.) * Claims 1-17 *	1-4,7, 8,13	
A	US-A-4 364 165 (J.E. DICKMAN et al.) * Claims 3,4; column 7, lines 41-59 *	1,3,8, 13	
A	US-A-4 356 042 (J.M. GEDALY et al.) * Claims 1-4 *	1,3,8	TECHNICAL FIELDS SEARCHED (Int. Cl. ³) H 01 L

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	10-10-1984	ZOLLFRANK G.O.

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO Form 1503. 03.82